

CLAIMS

1. A method of classifying input text to a target classification system having two or more target classes, the method comprising:

for each target class:

- providing at least first and second class-specific weights and a class-specific decision threshold;
- using at least first and second classification methods to determine respective first and second scores based on the input text and the target class;
- determining a composite score based on the first score scaled by the first class-specific weight for the class and the second score scaled by the second class-specific weight for the target class; and
- classifying or recommending classification of the input text to the target class based on the composite score and the class-specific decision threshold.

2. The method of claim 1, wherein at least one of the first and second scores is based on a set of one or more noun-words pairs associated with the input text and a set of one or more noun-word pairs associated with the target class, with at least one noun-word pair in each set including a noun and a non-adjacent word.

3. The method of claim 1, wherein providing each first and second class-specific weight and class-specific decision threshold comprises searching for a combination of first and second class-specific weights and class-specific decision threshold that yield a predetermined level of precision at a predetermined level of recall based on text classified to the target classification system.

4. The method of claim 1, wherein a non-target classification system includes two or more non-target classes, and at least one of the first and second scores is based on one or more of the non-target classes that are associated with the input text and one or more of the non-target classes that are associated with the target class.

5. The method of claim 4:

- wherein the input text is a headnote for a legal document; and
- wherein the target classification system and the non-target classification system are legal classification systems.

6. The method of claim 1, wherein the target classification system includes over 1000 target classes.

7. The method of claim 1, further comprising:

- displaying a graphical user interface including first and second regions, with the first region displaying or identifying at least a portion of the input text and the second region displaying information regarding the target classification system and at least one target class for which the input text was recommended for classification; and
- displaying a selectable feature on the graphical user interface, wherein selecting the feature initiates classification of the input text to the one target class.

8. A machine-readable medium comprising instructions for implementing the method of claim 1.

9. A method of classifying input text to a target classification system having two or more target classes, the method comprising:
for each target class:

- determining first and second scores based on the input text and the target class;
- determining a composite score based on the first score scaled by a first class-specific weight for the target class and the second score scaled by a second class-specific weight for the target class; and
- determining whether to identify the input text for classification to the target class based on the composite score and a class-specific decision threshold for the target class.

10. The method of claim 9, wherein at least one of the first and second scores is based on a set of one or more noun-words pairs associated with the input text and a set of one or more noun-word pairs associated with the target class, with at least one noun-word pair in each set including a noun and a non-adjacent word.

11. The method of claim 9, wherein determining the first and second scores comprises determining any two of:

- a score based on similarity of at least one or more portions of the input text to text associated with the target class;
- a score based on similarity of a set of one or more non-target classes associated with the input text and a set of one or more non-target classes associated with the target class;
- a score based on probability of the target class given a set of one or more non-target classes associated with the input text; and
- a score based on probability of the target class given at least a portion of the input text.

12. The method of claim 11, wherein each target class is a document and the text associated with the target class comprises text of the document or text of another document associated with the target class.

13. The method of claim 9:

- wherein determining the first and second scores for each target class comprises:
 - determining the first score based on similarity of at least one or more portions of the input text to text associated with the target class; and
 - determining the second score based on similarity of a set of one or more non-target classes associated with the input text and a set of one or more non-target classes associated with the target class;
- wherein the method further comprises determining for each target class:
 - a third score based on probability of the target class given a set of one or more non-target classes associated with the input text; and

- a fourth score based on probability of the target class given at least a portion of the input text; and
- wherein the composite score is further based on the third score scaled by a third class-specific weight for the target class and the fourth score scaled by a fourth class-specific weight for the target class.

14. The method of claim 9:

- wherein the input text is associated with first meta-data and each target class is associated with second meta-data; and
- wherein at least one of the first and second scores is based on the first meta-data and the second meta-data.

15. The method of claim 14, wherein the first meta-data comprises a first set of non-target classes that are associated with the input text and the second meta-data comprises a second set of non-target classes that are associated with the target class.

16. A machine-readable medium comprising instructions for performing the method of claim 9.

17. A system for classifying input text to a target classification system having two or more target classes, the system comprising:

- means for determining for each of the target classes at least first and second scores based on the input text and the target class;
- means for determining for each of the target classes a corresponding composite score based on the first score scaled by a first class-specific weight for the target class and the second score scaled by a second class-specific weight for the target class; and
- means for determining for each of the target classes whether to classify or recommend classification of the input text to the target class based on the corresponding composite score and a class-specific decision threshold for the target class.

18. A method of classifying input text according to a target classification system having two or more target classes, the method comprising:

- for each target class, determining a composite score based on a first score scaled by a first class-specific weight for the target class and a second score scaled by a second class-specific weight for the target class, with the first and second scores based on an input text and text associated with the target class; and
- for each target class, classifying or recommending classification of the input text to the target class based on the composite score and a class-specific decision threshold for the target class.

19. The method of claim 18, wherein the first and second scores are selected from the group consisting of:

- a score based on similarity of at least one or more portions of the input text to text associated with the target class;
- a score based on similarity of a set of one or more non-target classes associated with the input text and a set of one or more non-target classes associated with the target class;
- a score based on probability of the target class given a set of one or more non-target classes associated with the input text; and
- a score based on probability of the target class given at least a portion of the input text.

20. The method of claim 18, further comprising:

updating the class-specific threshold for one of the target classes based on acceptance or rejection of recommended classifications of the input text.

21. A method of classifying text to one or more target classes in a target classification system, the method comprising:

- identifying one or more noun-word pairs in a portion of text.

22. The method of claim 21, wherein identifying one or more noun-word pairs in the portion of text, comprises:

- identifying a first noun in the portion of text; and
- identifying one or more words within a predetermined numbers of words of the first noun.

23. The method of claim 21, wherein identifying one or more words within a predetermined number of words of the first noun comprises excluding a set of one or more stop words.

24. The method of claim 21, wherein the portion of text is a paragraph.

25. The method of claim 21, further comprising:

determining one or more scores based on frequencies of one or more of the identified noun-word pairs in the portion of text and one or more noun-word pairs in text associated with one of the target classes.

26. The method of claim 25, wherein the one or more scores include:

- at least one score based on similarity of at least one or more portions of the input text to text associated with the target class;
- at least one score based on similarity of a set of one or more non-target classes associated with the input text and a set of one or more non-target classes associated with the target class;
- at least one score based on probability of the target class given a set of one or more non-target classes associated with the input text; and
- at least one score based on probability of the target class given at least a portion of the input text.

27. The method of claim 25, wherein determining one or more scores based on one or more identified noun-word pairs and one or more noun-word pairs in other text associated with one of the target classes, comprises:

- determining a respective weight for each identified noun-word pair, with the respective weight based on a product of a term frequency of the identified word-noun pair in the text and an inverse document frequency of the noun-word pairs in the other text associated with one of the target classes.

28. A method of classifying input text to one or more target classes in a target classification system, the method comprising:

- identifying a first set of noun-word pairs in the input text, with the first set including at least one noun-word pair formed from a noun and non-adjacent word in the input text;
- identifying two or more second sets of noun-word pairs, with each second set including at least one noun-word pair formed from a noun and non-adjacent word in text associated with a respective one of the target classes;
- determining a set of scores based on the first and second sets of noun-word pairs; and
- classifying or recommending classification of the input text to one or more of the target classes based on the set of scores